American Journal of Medical Genetics (Neuropsychiatric Genetics) 74:112–114 (1997)

# *Rapid Publication*

COMMENTARY

## The Role of Meta-Analysis in Linkage Studies of Complex Traits

**John P. Rice**
*Department of Psychiatry, Washington University School of Medicine*
*St. Louis, MO, USA*

### INTRODUCTION

The analyses done by the Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 , and 8 are both ambitious and enlightening. The effort required to assemble data from 14 international research groups and code them for a combined analysis can not be overestimated. Moreover, the cooperation of the primary data contributors and the coordination of the data analysts are impressive. However, with this large sample, "the results are interpreted as inconclusive but suggestive of linkage" on chromosomes 6 and 8. The methods used were innovative in several respect, but also indicate the need for further work in combining linkage results from multiple studies.

We discuss one approach guided by the technique of meta-analysis (Hedges and Olkin, 1985) used to integrate results from multiple studies in the social sciences. Classical statistics primarily addresses the analysis of single experiments, and meta-analysis has emerged to summarize research from multiple independent experiments. As discussed below, these methods were largely irrelevant for the detection of linkage for homogenous Mendelian disorders, but may be important for synthesizing results for disorders such as schizophrenia.

When comparing results from multiple studies, the estimates of parameters must be compared, not levels of significance. For example, in a nongenetic study, the relationship between two continuous random variables may be of interest. One study may estimate a correlation coefficient of 0.2 which is highly significant, and a second study may yield an estimate of 0.3 which is not significant due to a small sample size. The second study should not be viewed as a non-replication; rather, a formal test of heterogeneity may be performed to test whether the two estimates differ. In this setting, a combined estimate may be easily derived.

Geneticists have traditionally used the maximum lod score Z to communicate results. However, the lod score is in fact a measure of statistical significance (Ott, 1995). When the mode of inheritance is known (so that only the recombination fraction $\theta$ is estimated), then 4.6 Z is asymptotically a chi-square with one degree of freedom. For a critical value of Z = 3, the corresponding chi-square value is 13.8 with a p-value between 0.001 and 0.001.

### METHODS

The central issue is how to integrate independent studies. As is the case in the 14 schizophrenia studies, samples come from different populations, different diagnostic schema are used, different sets of markers are typed in different laboratories, and different ascertainment procedures are followed in the individual studies.

Hedges and Olkin point out the fallacy of using outcomes of significance (i.e, in our case, lod scores). to interpret consistency and replication. These "vote-counting" methods give misleading conclusions. When comparing two studies, significance/non-significance may reflect design or power difference. Even when both studies are significant, there may be heterogeneity between the studies. If the maximal lod scores in two studies are 10 or 50 cM apart , are they consistent? There is no easy way to tell whether the results of two studies are consistent with one another from the outcome of their individual significance tests alone.

#### Genetic Effect Size

In meta-analysis, the effect size is the standardized difference in the means between experimental and control groups. This framework is often used in combining results from clinical trials. In traditional linkage studies, the recombination fraction $\theta$ is the key parameter. The lod curve $Z(\theta) = \log_{10} [L(\theta)/L(\theta = \frac{1}{2})]$, proportional to the likelihood ratio statistic, is computed and the lod score Z is the maximum value of $Z(\theta)$. For multiple studies, the individual lod curves may be added and a pooled $\theta$ estimated and tests of heterogeneity performed (Ott, 1995).

For a complex trait such as schizophrenia where the mode of inheritance is unknown, the use of $\theta$ is

Received 7 January 1997; Revised 8 January 1997

problematic. The recurrence risks for schizophrenia suggest an oligogenic, rather than a single locus trait (Risch, 1990). Programs such as LINKAGE (Lathrop et al., 1984) assume a single locus trait model, and specification of an incorrect trait model will bias the estimate of $\theta$. This will be especially problematic for multipoint linkage analysis. Thus, although $\theta$ is the natural parameter for single locus traits with known penetrances, it is probably not useful for a trait such as schizophrenia.

Recent analytic strategies have focused on affected relative methods. With respect to a single parent, a pair of siblings may either share or not-share the marker alleles transmitted by that parent. In general, we expect there to be 50% sharing with a parent. If we sample affected sib pairs, then we expect the sharing to be in excess of 50% if that marker is linked to a susceptibility locus for the disorder. For a (rare) recessive disease, sharing is 100% in affected sib pairs, and for a (rare) dominant disease) the sharing is expected to be 75%.

Consider the random variable S defined on affected sib pairs. S = 1 if the pair inherits the same allele from a particular parent, and S = 0 if not. When there is full information (i.e. the parents are fully informative at the marker) and there is one affected sib pair per family, then S is a simple binomial random variable.

Recent programs such as SIBPAL (SAGE, 1994), ASPEX (Hinds and Rich, unpublished) or MAPMAKER/SIBS (Kruglyak and Lander, 1995) are appropriate when parents are untyped. ASPEX and MAPMAKER/SIBS use unaffected individuals only for their genotypic information in performing multipoint analyses on affected sib pair. The program (GENEHUNTER (Kruglyak et al., 1996) allows for joint haplotype sharing in several affected relatives in a family. However, we will discuss only the statistic S. We refer to the mean of S (i.e. the percent sharing) S the genetic effect size.

## The Logistic Model for Haplotype Sharing

Since S is a dichotomous variable, we may model the effects of covariates using the logistic model. The logistic r of S is defined as $r = \log(S/(1-S))$, and we assume that $\Omega$ has a linear regression o covariates. For example, we may code the variable Par as Par = 1 for the mother and Par = 0 for the father. Then the model

$$\Omega = \alpha + \beta \, \mathrm{Par}$$

allows for a parent-of-origin effect, i.e. the maternal sharing may differ from the paternal sharing. Similarly, the sharing may be higher for sib pairs affected with a narrow diagnostic classification. For K studies, K - 1 "dummy" variables $X_1 \ldots, X_{K-1}$ may be defined with $X_i = 1$ for study I, and 0 otherwise. Then the model

$$\Omega = \alpha + \beta_1 X_1 + \ldots + \beta_{K-1} X_{K-1} + \beta_K \, \mathrm{Par} + \beta_{K+1} \, D$$

may be used to test for heterogeneity, parent-of-origin effect and a diagnostic affect D.

The above is meant to illustrate one approach for performing a meta-analysis on several studies by introducing covariates which can then be formally tested. The covariates may be polychotomous or ordinal and can be chosen to capture systematic differences between studies.

## Potential Covariates

The analysis by the Schizophrenia Linkage Collaborative Group indicates several covariates of potential importance.

- Diagnostic Criteria—Some of the studies used a narrow classification, some used a broad one (like the original positive report). The analyses focused on the narrow diagnosis available in all studies, and performed a separate analysis to examine the broad one of Straub and colleagues (1995). However, the article discussed lod-scores, rather than some genetic effect size, when considering the impact of diagnoses.

- Marker Allele Frequencies—When parents are not genotyped, the lod score (or estimated sharing) depends on the marker allele frequencies. Misspecification of these frequencies, or heterogeneity within the sample, can lead to spurious linkage evidence. For example, if 50% of the families in a population have allele "1" and 50% have allele "2", then all sib pairs would have identity by state of 2, with an allele frequency of 0.5 in the population, so that 50% of the parents would be estimated (falsely) to be heterozygotes and informative for linkage.

  In the Collaborative analysis, entirely different allele sets were assumed in each sample, so this potential problem was avoided. Indeed, without appropriate laboratory controls, it would be difficult to callibrate specific alleles across different labs.

- Ascertainment Schemes—The different studies used different criteria to sample and extend their families. It is possible that this could influence the degree of sharing seen in the individual samples. This was not investigated in their analyses.

- New vs. Original Reports—The original positive reports were separated from the new data in the Collaborative analyses. This is important since these regions are being studied due to these reports. From the meta-analysis perspective by including this as a covariate, we could also test for differences in the level of sharing and perform a formal test of heterogeneity between the original and new samples.

## DISCUSSION

The above comments are meant to illustrate an approach to combining multiple linkage studies, rather than a potential solution. It is clear that much can be gained from consideration of some type of genetic effect size rather than consideration of statistical significance. The percent sharing among affected sib pairs seems one logical candidate.

One lesson from other disciplines is that a single study (or set of studies with a common design) is preferable to *post hoc* analysis of a set of small studies. Often , the design differences can not be quantified, and the use of many covariates reduces the overall power of the combined data. If our field could agree on a consistent set of covariates, then investigators would be free to conduct studies independently, but also allow for future meta-analysis.

The Collaborative analyses indicate the importance of having access to the raw data coded in a similar format. This is a landmark study in this regard. The difficulty in interpreting the final answer to the basic question—Is there a schizophrenia susceptibility gene on chromosomes 6 or 8?—indicates that new methods are needed.

## ACKNOWLEDGMENTS

Supported in part by NIMH Grants MH37685, and MH31302.

## REFERENCES

Hedges LV and Olkin I (1985) "Statistical Methods for Meta-Analysis." San Diego: Academic Press, Inc.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996): Parametric and nonparametric linkage analysis: a unified multipoint approach. American Journal of Human Genetics. 58:1347-63

Kruglyak L, Lander ES (1995): Complete multipoint sib-pair analysis of qualitative and quantitative traits. American Journal of Human Genetics 57:439-54.

Lathrop GM, Lalouel JM, Julier C and Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443-3446.

Ott J (1991): "Analysis of Human Genetic Linkage." Baltimore : Johns Hopkins Press.

Risch N (1990) Linkage strategies for genetically complex traits: I. Multilocus models. American Journal of Human Genetics 46:222-228.

S.A.G.E. (1994): Statistical Analysis for Genetic Epidemiology, Release 2.2. computer program package available from the Department of Biometry and Genetics, LSU Medical Center, New Orleans.

Schizophrenia Linkage Collaborative Group for Chromosomes 3,6,and 8 (1996): Additional support for schizophrenia linkage on chromosomes 6 and 8: a multicenter study. American Journal of Medical Genetics (Neuropsychiatric Genetics) 67-580-594.

Straub RE, MacLean CJ, O'Neill FA, et al. (1995): A potential vulnerability locus for schizophrenia on chromosome 6p2422: evidence for genetic heterogeneity. Nature Genetics 11:287-93.